

Categorizing Web Pages and Data Mining

Zohreh vahedi*

Islamic Azad university of ferdows, Department of Computer engineering, Ferdows, South Khorasan, Iran

Abstract

The diversity of knowledge on the web has made determining communication patterns in the database and knowledge discovery among this mass of information an attractive target. The first step to achieve this goal is to classify web pages. The current machine learning techniques to classify content are initially discussed by flat and simple text documents and do not use structures such as links and headers discussed in the web optimally. Data mining is a set of techniques that allows a person to move beyond conventional data processing and help mining the information hidden in a massive volume of data. Today, in most organizations, data are collected and stored rapidly. However, using these data is not simple and they cannot be used as a single unit of the volume of data, so techniques they can be applied properly using a combination of statistics and computer science and the use of machine learning. However, in order to achieve a meaningful result of Web mining we need to have good data on our website, so effective management of Web data is crucial in web mining. The ultimate goal of this descriptive paper is that the organizations and small businesses could use data mining in their decisions and e-commerce as a future big business would enter into new future areas. There is a wide range of users considered for this technology, including for example research centers, companies operating in the field of web and database, analysts and managers of organizations, business, Web (search engines) and ...

Keywords: data mining, web mining, classification, e-commerce